You are comparing two linear models, m_1 and m_2, and are told that the first model has an R^2 of 0.815, and the second model has an R^2 of 0.765. Which of the two models do you prefer?

- a. There is insufficient information to prefer either model ✔

  {R^2} alone is not enough to prefer one model over another, for example, if you find out that the first model has 10,000 covariates, whilst the other has only 6 covariates, then the {R^2} differential does not seem very impressive. Recall that {R^2} tends to increase as new variables are included in the model, even if those variables are useless.
- b. m_2
- c. m_1
- d. Both models are equivalent

Which of the following is **not** an assumption of the univariate linear regression model with feature vector x, target output y, and predictions $\hat{y}$?

I. y is normally distributed for any fixed value of x.

II. x has a linear relationship with y

III. The variance of the residual $e = y - \hat{y}$ is constant for any value x

IV. All observations (x_i, y_i) are independent of each other

- a. I
- b. II
- c. III
- d. IV

Compute the derivative of $f(x) = x^3 ln(x) + \frac{3}{x^3}$.

- a. $f'(x) = 3x^2 \ln(x) + x^2 - \frac{3}{x^4}$
- b. $f'(x) = 3x^2 - \frac{9}{x^4}$
- c. $f'(x) = 3x^2 \ln(x) + x^2 - \frac{9}{x^4}$
- d. $f'(x) = 6x^2 \ln(x) + x^2 - \frac{9}{x^4}$

Compute the derivative of $f(x) = (x-4)(2x+x^2)^2$

- a. $f'(x) = 5x^4 - 18x^2 - 32x$
- b. $f'(x) = x^4 - 36x^2 - 16x$
- c. $f'(x) = x^4 - 36x^2 - 32x$
- d. $f'(x) = 5x^4 - 36x^2 - 32x$

Consider the following 3 datasets:

Dataset 1	
x	y
1	2
2	3
3	6
4	8
5	11

Dataset 2	
x	y
1	3
2	3
3	3
4	3
5	3

Dataset 3	
x	y
1	3
2	2
3	2
4	1
5	-4

Which of the following most accurately describes the correlations r_1, r_2, r_3 between x and y across datasets $1, 2, 3$ respectively?

- a. $r_1 > 0, r_2 = 0, r_3 > 0$
- b. $r_1 > 0, r_2 = 0, r_3$
- c. $r_1 > 0, r_2 = \text{undefined}, r_3$
- d. $r_1 > 0, r_2 = \text{undefined}, r_3 > 0$

Compute the derivative of $f(x) = \frac{\ln x+1}{x^2}$

- a. $f'(x) = -\frac{\frac{1}{x}}{2x}$
- b. $f'(x) = -\frac{2\ln(x)+1}{x^3}$
- c. $f'(x) = -\frac{2\ln(x)+1}{x^4}$
- d. $f'(x) = \frac{\frac{1}{x}}{2x}$

A new COVID19 home-testing kit is developed and administered to 100 individuals who are known to have the virus. Out of 100, the test returns 72 positive results.

- a. The test has precision of 72%
- b. The test has recall of 28%
- c. The test has recall of 28%
- d. The test has recall of 72%

Consider the following data: $X = ((1,1),(3,1),(1,-1),(2,-2))$ with corresponding labels $y = (1,1,-1,-1)$. The basic linear classifier is described by the equation $\langle x, w \rangle = t$ where:

- a. $w = (5/2, 1/2)$ and $t = 1/8$
- b. $w = (5/2, -1/2)$ and $t = -1/4$
- c. $w = (1/2, 5/2)$ and $t = 1/4$
- d. $w = (-5/2, 1/2)$ and $t = -1/4$

Which of the following is **not** an assumption of the univariate linear regression model with feature vector x, target output y, and predictions $\hat{y}$?

I. y is normally distributed for any fixed value of x.

II. x has a linear relationship with y

III. The variance of the residual $e = y - \hat{y}$ is constant for any value x

IV. All observations (x_i, y_i) are independent of each other

- a. I
- b. II
- c. III
- d. IV

Assume we have two classes A and B, and we wish to classify a new document d. We have the following training data:

document	class	cosine distance to d
1	A	1
2	B	0.95
3	B	0.94
4	A	0.45
5	A	0.4
6	B	0.39

Where we are using the cosine distance as a measure of distance (the higher the value, the closer the two documents). What class would be assigned to d using a k-NN classifier where we take $k = 3$ then $k = 5$.

- a. Class A for both $k = 3$ and $k = 5$
- b. Class A for $k = 3$ and Class B for $k = 5$
- c. Class B for both $k = 3$ and $k = 5$
- d. Class B for $k = 3$ and Class A for $k = 5$ ✔